

RESEARCH ARTICLE

Collaborative Fault Tolerance for Cyber-Physical Systems: The Diagnosis Stage

LUIS PIARDI^{1,2,3}, PEDRO COSTA^{1,3,4}, ANDRÉ SCHNEIDER DE OLIVEIRA¹, (Member, IEEE), AND PAULO LEITÃO^{1,2}, (Senior Member, IEEE)

¹Graduate Program in Electrical and Computer Engineering (CPGEI), Universidade Tecnológica Federal do Paraná-UTFPR, Curitiba, Paraná 80230-901, Brazil

²CeDRI, SusTEC, Instituto Politécnico de Bragança, 5300-253 Bragança, Portugal

³Faculty of Engineering, University of Porto (FEUP), 4200-465 Porto, Portugal

⁴Institute for Systems and Computer Engineering, Technology, and Science (INESC TEC), 4200-465 Porto, Portugal

Corresponding author: Luis Piardi (luispiardi@alunos.utfpr.edu.br)

This work was supported by the Coordination for the Improvement of Higher Education Personnel-CAPES under Grant ROR identifier: 00x0ma614 for the Article Processing Charge for the publication of this research. This work was supported in part by Fundação para a Ciência e a Tecnologia (FCT) through Portugal National funds FCT/MCTES (PIDDAC): CeDRI under Grant UIDB/05757/2020 (DOI: 10.54499/UIDB/05757/2020) and Grant UIDP/05757/2020 (DOI: 10.54499/UIDP/05757/2020), and in part by SusTEC under Grant LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020). The work of Luis Piardi was supported by FCT for the Ph.D. under Grant UI/BD/151286/2021 (DOI: 10.54499/UI/BD/151286/2021).

ABSTRACT The reliability and robustness of cyber-physical systems (CPS) are critical aspects of the current industrial landscape. The high level of autonomous and distributed components associated with a large number of devices makes CPS prone to faults. Despite their importance and benefits, traditional fault tolerance methodologies, namely local and/or centralized, often overlook the potential benefits of collaboration between cyber-physical components. This paper introduces a collaborative fault diagnosis methodology for CPS, integrating self-fault diagnosis capabilities in agents and leveraging collaborative behavior to enhance fault diagnosis. The contribution of this paper relay in propose a methodology for fault diagnosis for CPS, based on multi-agent system (MAS) technology as a backbone of infra-structure, highlighting the components, agent behavior, functionalities, and interaction protocols, to explore the benefits of communication and collaboration between agents. The proposed methodology enhance the accuracy of fault diagnosis when compared with local approach. A case study was conducted in a laboratory-scale warehouse, focusing on diagnosing drift, bias, and precision faults in temperature and humidity sensors. Experimental results reveal that the collaborative methodology significantly outperforms the local approach in fault diagnosis, as evidenced by performance improvements in diagnosis classification. The statistical significance of these results was validated using the Wilcoxon signed-ranks test for paired samples.

INDEX TERMS Fault diagnosis, cyber-physical system, multi-agent system, collaboration, fault tolerance.

I. INTRODUCTION

The communication, information exchange, and collaboration among individuals of a species (e.g., humans or ants) are fundamental to enhancing knowledge, resilience in crises, and problem-solving to ensure species survival. By combining the skills and knowledge of different individuals, this approach enables a more efficient resolution of complex problems with a wider set of information. An analogy between these species and digital systems can be established,

The associate editor coordinating the review of this manuscript and approving it for publication was Guillermo Valencia-Palomo¹.

where communication, information exchange, and collaboration transcend the individual capacity of intelligent and autonomous components. Thus, in addition to performing their functions autonomously, the component's interaction, data exchange, and information translates into continuous improvement, adapting to different operating conditions culminating in collective intelligence [1].

To foster the collaboration among the components, the information and communication technology (ICT) embedded in cyber-physical systems (CPS) plays a fundamental role in digitization. This procedure involves the integration of technologies such as artificial intelligence (AI), Internet of

Things (IoT), and cloud computing, along with interoperability standards between different components [2]. CPS are complex engineered systems that monitor, coordinate, control, and integrate physical devices or processes using embedded computing, sensing, and network communication [3]. By integrating autonomous components and entities into a unified system through the tight interaction of cyber and physical domains, CPS provides a robust structure to meet the requirements of scalability, robustness, flexibility, and dependability while delivering advanced functionalities [4]. In this context, the collective intelligence of CPS emerges from the collaboration between components that exhibit local intelligence [5].

There is an increasing demand for reliability and robustness in CPS [6], as their components are constantly susceptible to anomalies and fault episodes [7]. Detecting and diagnosing potential faults at the earliest opportunity is crucial, as well as implementing fault-tolerant operations to minimize performance degradation and prevent hazardous conditions. Fault tolerance (FT) is essential for maintaining component operation during fault episodes. One of the key challenges in fault tolerance (FT) for cyber-physical systems lies in the lack of approaches that fully leverage CPS's intrinsic capabilities for communication, interaction, and autonomous decision-making dedicated explicitly to enhancing fault tolerance. Current approaches tend to rely on either centralized strategies, where a single node is responsible for FT across the system, leading to issues such as a single point of fault and limited scalability, or on local strategies, where each component independently manages its own faults, which problems related to restricted awareness and limited resources for implementing robust FT mechanisms impacting the accuracy of the results. This lack of collaborative FT methodologies, wherein components interact and share information to enhance FT, limits the CPS adaptability and resilience. An approach that exploits CPS's interconnected and interactive nature could significantly improve FT across all FT stages, namely fault detection, diagnosis, and recovery [8].

In this context, this paper presents the Collaborative Fault * (CF*), a distributed and collaborative fault tolerance methodology that integrates self-fault tolerance capabilities in cyber components and employs collaboration strategies to enhance component fault tolerance. Although CF* addresses the entire path of collaborative fault detection, collaborative fault diagnosis and collaborative fault recovery, this work focuses on the collaborative fault diagnosis stage. Multi-agent system (MAS) concepts are used to introduce autonomy, intelligence (for example, based on machine learning or deep learning algorithms), and communication into components, enabling exploration of the effects of collaboration during the fault diagnosis stage. The objective is to improve the fault diagnosis capability and accuracy through interaction, communication, and collaboration between agents, highlighting the agent's behavior, functionalities, and interaction protocols. The proposed approach was experimentally tested using

a case study based on a laboratory small-scale warehouse, with fault diagnosis tests conducted on the temperature and humidity sensors of the racks. Each agent, representing the cyber part of rack, has embedded machine learning models for conducting fault diagnosis. Validation was performed using the Wilcoxon signed-ranks test for paired samples, comparing the impact on fault diagnosis performance in terms of F1-score (a metric that balances precision and recall for fault classification models) for both local (non-collaborative) and collaborative CF* methodology. The preliminary results show the proposed methodology improves the diagnosis accuracy by up to 10% in terms of median compared to the local approach.

This work is organized as follows: the section II presents the related work on fault diagnosis in the context of CPS. The section III presents the proposed CF* methodology, emphasizing fault diagnosis behaviors and collaboration protocols. Then, section IV presents the case study, the preparation of the data set for fault diagnosis, and the experimental procedure. The experimental results with statistical tests are discussed in section V, and finally, section VI highlights the conclusions and presents future work.

II. RELATED WORK IN FAULT DIAGNOSIS FOR CPS

The advent of CPS has significantly changed the relationship between the logical and physical layers of digital systems. This transformation has been driven by the digitization of systems, with the intensive incorporation of technologies such as IoT, generating an ever-increasing volume of data [9]. Consequently, there has been a need to shift from the centralized paradigm to distributed and decentralized architectures for device control and condition monitoring to process and extract useful information from this large volume of data [10]. This new technological context aims to improve data processing, and analysis, making CPS more robust and reliable.

Despite this paradigm shift, studies and applications on fault tolerance in CPS are still supported by traditional methodologies [11]. In the centralized methodologies, digitized data is analyzed and processed by a central node, which uses methods, e.g., data-driven or model-based methods, for decision-making to execute the stages of fault detection, diagnosis, and recovery. In the local methodologies, each component has embedded algorithms for fault detection, diagnosis, and recovery. Both methodologies have disadvantages, for example, centralization can lead to single point of failure and overload the central node, while the local approach can result in inconsistent results due to myopia and lack of context, making the decision methods for the fault tolerance stages inefficient [11].

This paper focuses on the fault diagnosis stage, which consists of accurately identifying and classifying different types of faults. It allows the fault to be corrected quickly, mitigating its impact on the system and contributing to the fault tolerance process [12], [13]. In the context of CPS, fault diagnosis presents significant challenges with distributed

components. The components must have a uniform and comprehensive perception of the system, with the ability to diagnose and determine the magnitude and impact of the fault on the system. If faults are not appropriately diagnosed, they can quickly result in more critical problems, including catastrophic failures, fault propagation between components and sub-components, and an unreliable system state. Therefore, making the components communicate and learn from each other is necessary, promoting collaboration to make fault diagnosis more efficient in CPS.

Several recent studies explore and discuss fault diagnosis in CPS, adopting different approaches to improve the diagnosis regarding response time and fault classification performance in diverse applications. For example, the literature review presented in [14] focuses on fault diagnosis in CPS, specifically in the aerospace, automotive, and industrial control. This review analyzes the main fault diagnosis techniques, including Physics-based Models, Data-Driven Artificial Intelligence, and Knowledge-Based and hybrid techniques that combine aspects of these three approaches [14]. Another relevant work in the area reviews the state-of-the-art of machine learning-based fault diagnosis models across different domains, focusing on transfer learning strategies between components [15]. This review leverages models previously trained on several cyber components to diagnose similar faults, thus optimizing the diagnosis process [15].

With the expansion of IoT digitization and AI techniques, new machine-learning methods for fault diagnosis have rapidly advanced across various CPS applications. For instance, a Long Short-Term Memory algorithm has been proposed for fault diagnosis in wind turbines, effectively recognizing fault patterns in time series data acquired from sensors [16]. Another study employed the Support Vector Machine (SVM) algorithm to diagnose misalignment, unbalance, and backlash faults in induction motors using multiple vibration sensors [17]. Additionally, a deep perceptual adversarial domain adaptation algorithm for fault diagnosis in robotic manipulator joint bearings demonstrated superior performance compared to traditional convolutional neural networks under varied operational conditions [18].

In addition to the advances promoted by integrating AI in cyber components, information exchange and collaboration emerge as essential pillars to improve the effectiveness of fault diagnosis in CPS. Some studies have explored the collaboration, such as an investigation applied to the chemical process industry that uses a MAS framework to diagnose faults in distillation units, combining classifiers from different machine learning models to improve the fault diagnosis performance [19]. Another study presents a methodology that integrates multiple machine learning models for fault diagnosis, evaluating the results with the Tennessee Eastman benchmark dataset [20]. Additionally, collaboration is explored through the fusion of data from multiple sensors installed in a single asset, such as in measuring engine vibrations to diagnose faults [21]. Despite these innovative approaches, there are common limitations

that can be improved. These presented studies are based on the centralization of diagnostic decisions, which can result in a single point of failure and limitations on scalability.

To diagnose potential faults within CPS, a study proposed a composite hazard analysis technique that integrates content relationships among various hazard analysis methods, including ETA, FMEA, and FTA. This approach resulted in the development of CPSTracer, a tool designed to analyze potential hazards by generating a fault traceability graph that maps out fault diagnoses and fault propagation routes [22]. Another study in this area utilizes a resilient state transition diagram to model multiple failures within a platooning system, incorporating elements such as failure types, mitigation strategies, and safe exit procedures to achieve resilience in autonomous platooning contexts [23].

According to the review carried out, an existing gap was identified related to distributed methodologies for detailed, robust, and distributed agent collaboration in fault diagnosis for CPS applications, aiming to minimize the vulnerabilities and limitations of centralized and local approaches. The agent collaboration in fault diagnosis for CPS remains incipient, requiring more research and precise specifications to reach maturity and fully leverage its benefits in practical applications. This study addresses this gap by advancing the parameterization of agent behaviors and interaction protocols, enabling fault diagnosis through a distributed and collaborative methodology.

III. DESIGN OF COLLABORATIVE FAULT TOLERANCE: EMPHASIS ON DIAGNOSIS STAGE

The establishment of collaborative fault tolerance capability requires an environment consisting of a network of cyber-physical components, each endowed with local intelligence and compatible collaboration models. The interaction and data exchange between these distributed components enhance fault tolerance by enabling a holistic understanding of the system, thereby facilitating precise fault detection, diagnosis, and recovery. This section elucidates the proposed methodology for distributed and collaborative fault tolerance (CF*) in CPS, with an emphasis on diagnostic stage, detailing the agent's behavior and the main collaboration protocols.

A. OVERVIEW OF CF*

The CF* employs a multi-agent system concept to deploy collaborative functionalities in CPS components, as shown in Figure 1. Each physical component is paired with a cyber component (i.e. an agent) which, beyond behaviors of monitoring and control, can detect, diagnose and recovery from faults and collaborate to assist other agents in FT procedures.

The fault tolerance capability of the CF* methodology manifests on two levels: local and collaborative as shown in Figure 2. Each agent possesses distinct models for detecting, diagnosing, and recovering from failures, with unique features arising from their training datasets and model specifications. When a fault is well known and aligns with

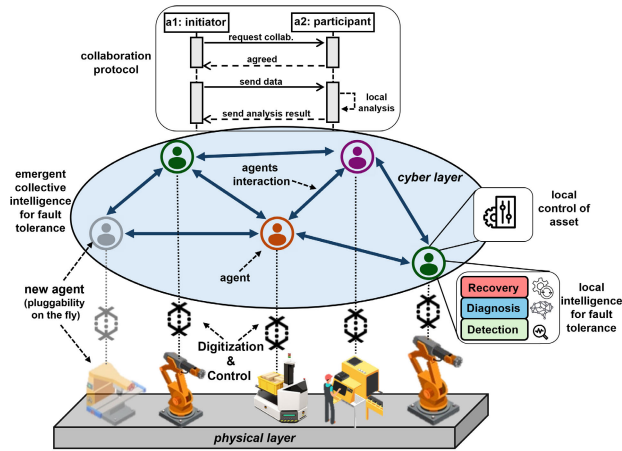


FIGURE 1. Representation of MAS based CPS.

the agent’s detection, diagnosis, and recovery models, the local level is sufficient for fault tolerance. However, when the agent’s models cannot address the fault independently, the collaborative level is activated, utilizing the experience and information from compatible agents to find a possible solution for the issue.

At the local level, all stages occur within individual agents, which execute self-fault tolerance processes to respond quickly to disturbances and minimize the fault impact. For instance, an agent responsible for a robot can detect a disturbance in movement speed, then diagnose the issue as a problem with the controller gains, and recover by adjusting these gains to achieve the desired speed and movement for the robot operation.

The collaborative level of fault tolerance is activated when any agent cannot perform one or more fault tolerance stages individually. In such situation, this agent assume the role of initiator, thereby instigating a structured collaboration process for the fault stage, wherein the initiator agent request assistance of other compatible agents based on the needed stage. For example, if uncertainty arises during the detection stage, the initiator agent collaborates with other agents’ detection stages. This principle applies similarly to the diagnostic and recovery stages. Multiple stages may require collaboration for the same event, for instance, detection might be handled locally, while diagnosis and recovery involve collaboration. Decisions regarding detection, diagnosis, or recovery are facilitated through agent interaction, enabled by collaboration and data exchange.

In the CPS context, a fault can affect one or more cyber-physical components, resulting in system performance degradation when faults propagate over the physical components. To determine whether the fault impacts only the local agent or the broader system (i.e., multiple agents), the diagnostic stage transmits an alert to other agents’ detection stages after diagnosing the fault. This procedure (illustrated by the blue arrows in Figure 2) allows other agents to assess whether the fault could affect them.

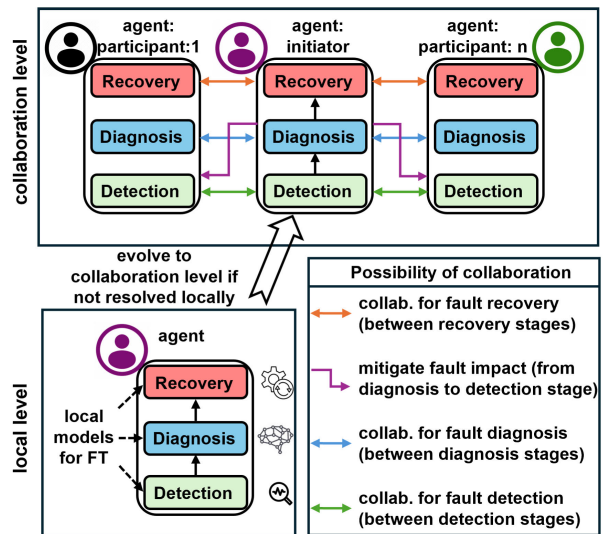


FIGURE 2. Local and collaborative fault tolerance levels. If the agent cannot tolerate the fault individually, it evolve for a collaborative FT.

The emergent collaborative FT behavior is a consequence of the synergistic collaboration among autonomous agents, wherein decision-making processes are decentralized. This approach addresses the classic issues associated with traditional local and monolithic fault tolerance methodologies, such as scalability limitations, system myopia, reduced performance, and single point of fault. A formalized approach related to the CF* methodology specifically for the detection stage can be consulted in [24].

B. COLLABORATION IN DIAGNOSIS STAGE

When a disturbance occurs in the system, the detection stage is responsible for detecting the fault as quickly as possible, preventing intensification and aggravation of the disturbance [7]. If fault detection is confirmed, the diagnosis stage comes into action. The functionality of this stage is to mitigate the cause and nature of the fault by providing information necessary for the system to elaborate a strategy for operating with or even recovering from the fault [12]. For instance, the diagnostics stage identifies which component, e.g., sensor or actuator, has failed, classifying the type and nature of the fault. Furthermore, it is essential to minimize fault misclassifications, which can lead to a mistaken recovery, further compromising the component’s operation.

For the collaborative fault diagnosis in CF*, consider a MAS with a set of j agents, where each agent a_i possesses local diagnosis models and capabilities. Initially, the diagnosis stage occurs locally, triggered when a fault is detected by the detection stage in the input sample ϵ_i , derived from digitized data related to the physical part of agent a_i . Each a_i employs a fault diagnosis method designed to handle k pre-defined fault classes. Consequently, the diagnosis objective of agent a_i is to categorize each input

sample (ϵ_i) detected as a fault into one of the k classes, yielding the diagnostic output $fdi_i(\epsilon_i)$ according to Eq. 1. The diagnostic model attempts to link each unusual or anomalous event observed in the agent a_i to its corresponding fault class.

$$fdi_i(\epsilon_i) = \begin{cases} 1, & \text{if fault type 1} \\ 2, & \text{if fault type 2} \\ \dots & \\ k, & \text{if fault type } k \\ k + 1, & \text{if fault unknown} \end{cases}, \quad \begin{cases} i \in \mathbb{N}^* \cap [1, j] \\ k \in \mathbb{N}^* \end{cases} \quad (1)$$

The k fault types represent known faults to which the system is susceptible. The diagnostic model correlates each anomalous behavior with the respective fault. This model can use different methods, such as model-based (based on the mathematical or physical model of the system), data-driven (AI models trained with historical data from the system) and knowledge-based (expert knowledge of the system encoded as rules heuristics). Furthermore, diagnostic models have an associated function $S(fdi_i(\epsilon_i))$ to measure the degree of similarity to the sample ϵ_i , measuring how closely the current fault event resembles or aligns with the known patterns associated with different fault classes. The similarity determines the confidence (or the probability) that the model attributes to each class of fault, according to Eq.2.

$$S(fdi_i(\epsilon_i)) = \begin{cases} s_1, & \text{similarity for fault type 1} \\ s_2, & \text{similarity for fault type 2} \\ \dots & \\ s_k, & \text{similarity for fault type } k \\ s_{k+1}, & \text{similarity for unknown fault} \end{cases}, \quad \begin{cases} x \in \mathbb{N}^* \cap [1, k + 1] \\ k \in \mathbb{N}^* \end{cases} \quad (2)$$

The similarity $S(fdi_i(\epsilon_i))$ indicates the confidence of the agent a_i to diagnose a fault. Likewise, the similarity may indicate that the diagnostic model does not recognize and cannot classify a certain event. The collaborative level starts case the data analyzed by the agent’s diagnostic model results in uncertainty in the similarity, meaning that there are no substantial confidence of model to classify the given input. Different diagnostic models may utilize differing metrics to quantify the degree of similarity $S(fdi_i(\epsilon_i))$. For instance, models that employ statistical pattern recognition for fault diagnosis may calculate similarity using metrics such as cosine similarity or Euclidean distance. Alternatively, if the diagnostic model is based on machine learning, similarity could be derived from a confidence score generated through probabilistic inference.

To carry out collaboration, the agent must meet three key parameters: determining when to initiate collaboration, identifying which agents to collaborate with, and analyzing the contributions of these agents to execute the fault diagnosis

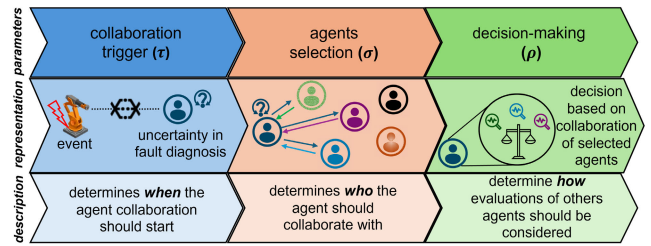


FIGURE 3. Procedure of collaborative fault diagnosis of CF*.

decision. The collaboration sequence is depicted in Figure 3, highlighting the collaboration trigger τ , agent selection σ , and decision-making ρ as the key collaboration parameters. By appropriately configuring these collaboration parameters, the agents can dynamically adapt their behavior to the specific needs and context of the CPS.

The collaboration trigger parameter τ indicates to the agent when is necessary to collaborate, i.e., the agent’s reasoning mechanism that informs when it is necessary to shift from a local to a collaborative level, representing that the similarity obtained for the diagnosis is lower than desired of others for the component. The agent selection parameter, σ , select the most suitable agents for collaboration, for example, considering the compatibility of diagnostic models of analogous and compatible agents. Finally, the decision-making parameter, ρ , indicate how the agent analyzing the contributions of other agents to execute a decision about the fault diagnosis. For example, decision-making techniques may include a weighted function based on the degree of analytical similarity from each selected agent or a majority voting system.

Table 1 presents examples of potential strategies that can be instantiated by system designers and engineers for each collaboration parameter, illustrating when, who, and how the collaboration process should occur. The correct instantiation of these parameters directly influences the collaborative behavior of each agent. Moreover, the flexibility of these strategies is a crucial factor, as they must be tailored to the specific characteristics of each agent and the particular context of the CPS application. Table 1 offers examples for each parameter that can be adapted and customized for different scenarios, thereby enhancing the effectiveness of the CF* approach. Furthermore, as detailed in [25], seven evaluation metrics are proposed to fine-tune these parameters during the agent operation, assessing potential issues such as excessive collaboration, agent hesitation, or errors resulting from collaboration. These evaluation metrics provide guidelines for refining these parameters to maximize the benefits of collaboration while mitigating unintended behaviors that could result from the initial parameter settings.

C. FAULT DIAGNOSIS COLLABORATION PROTOCOLS

Agents collaborate for fault diagnosis using interaction protocols conforming to the FIPA-ACL standards [26], such

TABLE 1. Examples of instantiation strategies for the collaborations parameters of CF*.

Parameter	Instantiation Strategies	Strategy Description and Examples
Trigger τ	uncertainty or inconclusiveness	this occurs when data analysis results in uncertainty for some stage of FT, e.g., low similarity in inference for making a decision in detection, diagnosis, or recovery.
	collaboration defined	collaboration pre-defined, e.g., time-out or after a certain amount of inferences.
	limited perception	case the agent's information is insufficient to make a decision, collaboration begins, for example, system information myopia.
	limited resources	this occurs when the agent faces a computational overhead, e.g., processing overhead for executing the detection, diagnosis, or recovery models.
Selection σ	proximity	selects compatible agents that are physically near.
	computational capacity	selects compatible agents with the greatest processing power.
	availability	selects compatible agents that are idle
	most voted	decisions based on the majority of votes obtained from other agents
Decision-Making ρ	weighted decision	decision based on the weighted votes of other agents
	more relevance	decision made by the agent most confident in the collaboration

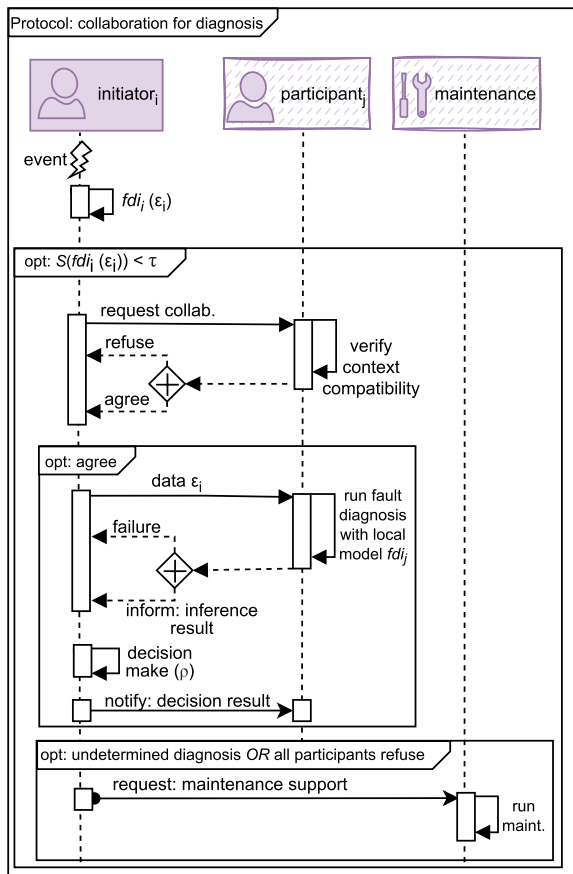


FIGURE 4. Collaborative fault diagnosis protocol.

as FIPA-request and FIPA-query. For instance, an agent can use FIPA-request to inquire about another agent’s operating status to determine its availability for collaboration. Beyond these conventional protocols, CF* introduces a customized collaboration protocol for fault diagnosis, shown in Figure 4.

The collaborative fault diagnosis protocol in Figure 4 extends the FIPA-request functionality to enable more sophisticated collaboration. After executing the diagnosis model $fdi_i(\epsilon_i)$ and activating the collaboration trigger τ , the initiator requests collaboration to compatible agents. The

initiator agent shares diagnostic data with all participant agents selected according to the σ parameter. Once responses are received from the participants, the initiator agent makes a decision using the ρ strategy. This decision is then shared with all participant agents, promoting transparency and collective awareness within the MAS, as well as providing useful information for updating each agent’s diagnostic models and collaboration parameters, thereby improving fault diagnosis capabilities over time.

In alignment with each agent’s autonomy, the initiator agent is responsible for determining the diagnosis by applying a decision-making (ρ) based on the inputs gathered from participating agents during collaboration. Different agents may have varied experiences and diagnostic models, and conflicting diagnoses may arise. In cases when the decision policy is insufficient to resolve these conflicts and ensure diagnostic consistency or collaboration is not achieved, whether due to refusal from all participant agents, the initiator agent will escalate the issue to external maintenance support to analyze the event and provide an appropriate diagnosis.

Additionally, CF* establishes a protocol to determine whether a fault diagnosed is local or global, which is critical for appropriate recovery actions. After diagnosing a fault, the initiator agent (the one identifying the fault) assesses the potential for the fault to impact other agents in the system, as illustrated in Figure 5. The initiator shares diagnostic information with the participant agents (i.e., potential agents who might be impacted by the failure), who independently evaluate whether the fault has an impact on their own operation. Each participant then reports back to the initiator with their assessment. If these responses confirm fault propagation across multiple agents, the initiator classifies the fault as global rather than local. This classification is essential for the subsequent fault recovery stage, as it indicates that collaborative actions may be required to avoid the fault’s broader effects and protect the system’s overall functionality.

IV. DESCRIPTION OF THE CASE STUDY AND EXPERIMENTAL PROCEDURE

To experimentally validate the benefits of collaboration in fault tolerance, specifically in the diagnostic stage, a case study was conducted in which CF* was based on data-driven methods, with agents instantiated using machine learning models for fault diagnosis. The remainder of this section describes the case study, the dataset elaboration, and the evaluation methodology proposed with deployment of CF* in the agents.

A. CASE STUDY

The collaborative fault diagnosis was applied in ARENA [27], a laboratory small-scale warehouse platform, illustrated in Figure 6. One of the ARENA’s asset classes is the racks. Each rack is linked to a corresponding agent with processing capabilities for autonomous operation and collaboration to achieve warehouse goals. Each rack incorporates a cyber component represented by an agent responsible for

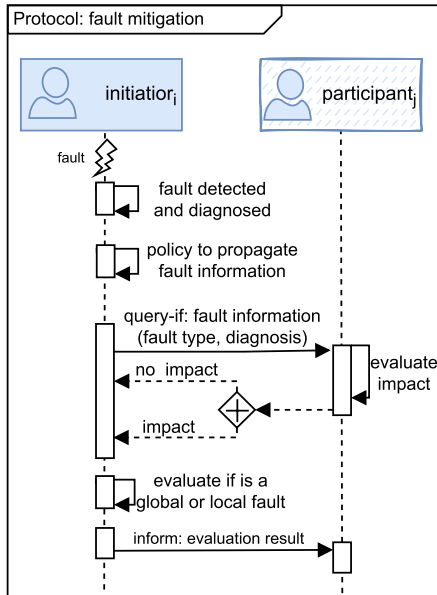


FIGURE 5. Fault mitigation protocol.

monitoring temperature and humidity values, which are pivotal for maintaining the desired storage conditions. Each agent is assigned to diagnose faults for its specific rack according to its fault diagnosis models. Adopting the CF* approach, these rack agents can engage in collaborative efforts with other rack agents to enhance diagnostic accuracy. Further details on the MAS-based CPS architecture implemented in the ARENA can be found in [27].

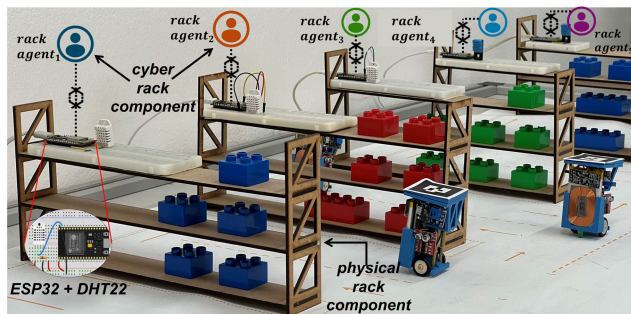


FIGURE 6. ARENA small-scale warehouse platform, highlighting the racks cyber physical components.

This case study focuses on the rack agent class to evaluate the impact of collaboration on fault diagnosis. The agents examine the condition of the physical rack component, where temperatures range from 17 to 20 °C and humidity levels vary between 35% and 40%. Any divergence from the expected sensor readings for temperature or humidity may compromise the quality and integrity of the stored items. Hence, ensuring the effectiveness of fault diagnosis for classifying the temperature and humidity irregularities is crucial, supporting a recovery stage that reduces the impact of any faults in the stored items. For instance,

suppose a rack’s temperature and humidity sensor system fails, potentially compromising the integrity of its products stored. The affected rack’s fault diagnosis model may lack the confidence to accurately diagnose the fault type and impact. To address this, the corresponding agent initiates collaboration with other rack agents to improve its diagnostic capabilities. This process involves triggering collaboration, selecting compatible agents to collaborate, exchanging fault data, allowing the compatible agents to perform inferences using their knowledge and diagnostic models. Subsequently, the corresponding agent formulates a diagnosis based on supplementary diagnostic data obtained from compatible agents.

Figure 6 illustrates the cyber-physical rack used in the experiments, which involved twelve agent racks. Each rack’s physical component consists of a wooden framework for item storage, equipped with a DHT22 sensor connected to an ESP32 microcontroller. This setup measures temperature and humidity, transmitting the data via Wi-Fi to the corresponding rack agent.

Each agent rack is deployed on a Raspberry Pi 3B using a proprietary framework coded in Python. The agents communicate using the collaborative fault diagnosis interaction protocols depicted in Figure 4. These protocols encode messages and data in JSON format, providing a simple and computationally efficient method for structuring information.

B. DATA SET ELABORATION

The temperature and humidity data from the ARENA racks were obtained using fourteen DHT22 sensors in a controlled room with temperatures ranging from 17 to 20 °C and humidity levels between 35% and 40%. Each temperature and humidity observation was recorded within a 10-second interval. Data from twelve sensors were used to train the twelve rack agents, each trained exclusively on data from its respective rack, resulting in distinct diagnostic models. The remaining two sensors were used to create two samples to test the diagnostic performance of each agent.

Since all the sensors used in the experiments were defect-free and the collected data had no faults, data processing was conducted on the ESP32 of each rack to inject three categories of faults into the sensor outputs. The fault categories are drift, bias, and precision, each exhibiting a trend that distinguishes it from expected sensor behavior. The work [28] inspired the fault injection models.

Under normal conditions, i.e., without faults, the DHT22 sensors provide temperature and humidity data according to mathematical model of Eq. 3.

$$o(t) = h(t) + \eta, \quad \eta \sim N(0, \delta_\eta^2) \tag{3}$$

where $o(t)$ is the (temperature or humidity) sensor output expected in time t of the sensor without the presence of a fault and $\eta \sim N(0, \delta_\eta^2)$ is the noise with zero mean and low variance associated with each sensor.

The drift fault has a characteristic behavior where its output increases continuously at a constant rate. This fault

is associated with corrosion of the sensor material over time [28]. The Eq. 4 presents the model for this fault.

$$o_{drift}(t) = h(t) + \eta + b(t), \quad b(t) = b(t - 1) + constant \quad (4)$$

where $b(t)$ is the drift added to the sensor output at time t , increasing over time at a constant rate, resulting in an increment in the sensor output.

Faults with characteristics of bias behavior have a constant value added to the sensor output, and, as a result, a shift from the expected sensor output is observed. The Eq. 5 expresses this characteristic of the bias fault. This behavior is associated with sensors with hardware problems, such as incorrect calibration or damaged components due to short circuits [28].

$$o_{bias}(t) = h(t) + \eta + constant \quad (5)$$

The third fault category is related to sensor precision. For this fault to occur, precision degradation is injected according to the Eq. 6. The disturbance related to this fault has zero mean and high variance.

$$o_{precision}(t) = h(t) + \eta + v, \quad v \sim N(0, \delta_v^2) : \delta_v^2 \gg \delta_\eta^2 \quad (6)$$

The precision fault occurs due to physical connection problems in the circuit, noise resulting from high-frequency procedures, and physical damage to the sensor [28].

Based on these fault models, a data set was composed of the three fault categories. Table 2 presents the data set distribution with the data divided between data training and data testing. In total, there are twelve training samples with drift, bias, and precision faults, and each sample is used to train the diagnostic models of a specific agent. The two remaining samples were used for experimental test to evaluate and compare the performance of the agent models with local or collaborative methodology for fault diagnosis.

TABLE 2. Training and test data for diagnosis for DHT22 sensor.

data set	Samples	Fault category	Amount of observations
data training	12 (one for each agent)	Drift	12000 (1000 for each sample)
		Bias	12000 (1000 for each sample)
		Precision	12000 (1000 for each sample)
data test	2 (one for each test)	Drift	1044 (522 for each sample)
		Bias	1162 (581 for each sample)
		Precision	1162 (581 for each sample)

Figure 7 presents the data test used for the experiments, which have the same faults but manifested with different intensities. Since the diagnosis is only performed after fault detection, the data used in the experiments corresponds only to the gray area of the data set (in accordance with Table 2).

C. EXPERIMENTAL PROCEDURE

The experiments were designed to evaluate the impact of collaboration on fault diagnosis, specifically whether it improves the ability to classify a fault category correctly. Therefore, each agent’s individual fault diagnosis stage (i.e.,

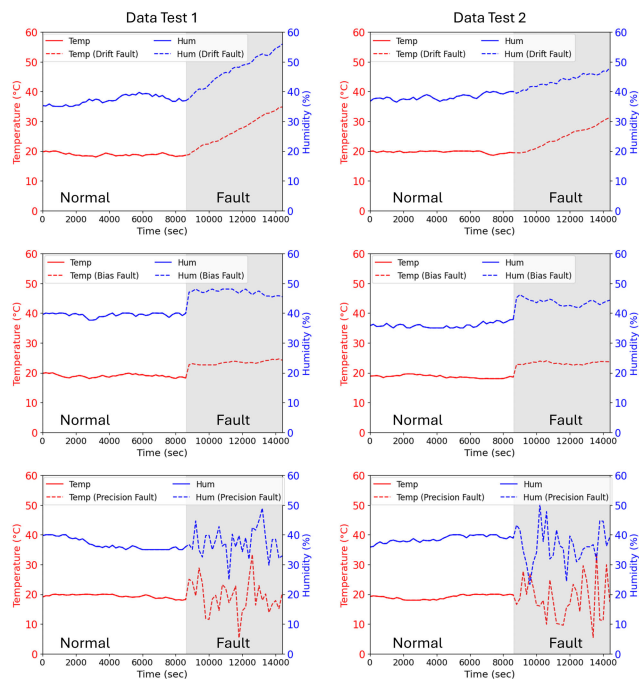


FIGURE 7. Data test with faults highlighted in the gray area.

local methodology) was compared with the collaborative fault diagnosis based on the CF* specifications adopting the fault diagnosis protocol, as depicted in the flowchart of Figure 8.

The data-driven method based on machine learning was adopted for the diagnosis due to its strong generalization capability for diagnosing faults in temperature and humidity sensors. Five machine learning algorithms were used in the experiments: Support Vector Machine (SVM), Nu-Support Vector Machine (NuSVM), Multi-Layer Perceptron (MLP), Random Forest (RF), and Gaussian Naive Bayes (GNB), to assess the impact of collaboration on different algorithms (evaluated with two distinct data test). These algorithms were chosen because they are widely used and representative in the context of fault diagnosis, as demonstrated in studies such as [29] and [30].

The diagnosis models ($f_{di}(\epsilon_i)$) were implemented in each agent a_i using the Python library scikit-learn [31]. For all models, the inputs (ϵ_i) consist of a time window containing sixty temperature and humidity observations. The output categories were drift, bias, or precision. All models (SVM, NuSVM, MLP, RF, and GNB) were trained with their respective default parameters according to the scikit-learn library. To obtain the similarity $S(f_{di}(\epsilon_i))$ of each classification, the inherent probability of each model was used, calibrated with the Platt Scaling method, available in the scikit-learn library, to obtain a more accurate probability [32].

To deploy the CF*, each agent performed self-fault diagnosis and collaboration as needed, based on the collaboration parameters τ , σ , and ρ . These parameters were applied uniformly across all agents. Collaboration was triggered

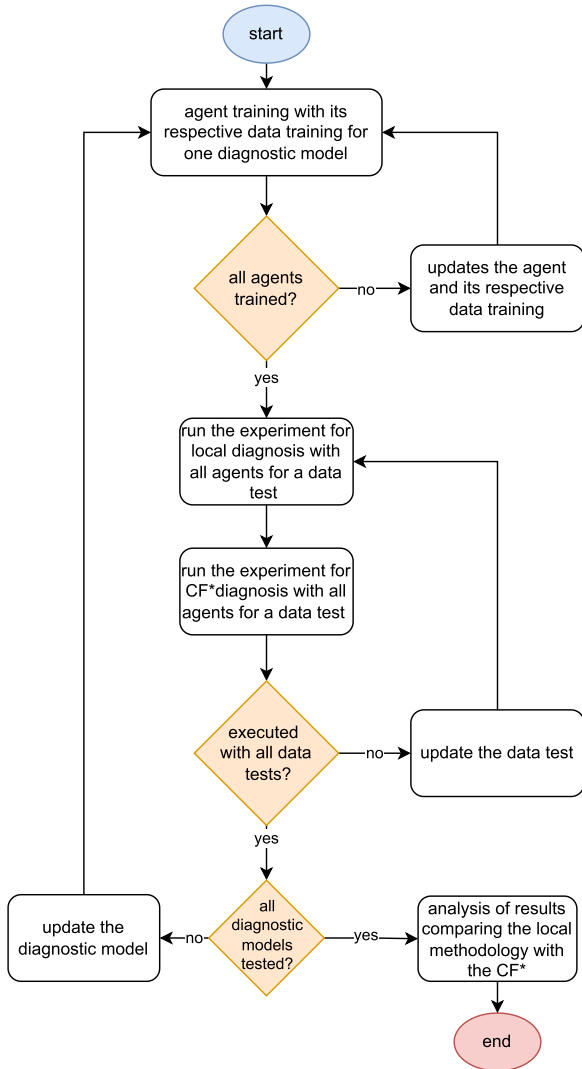


FIGURE 8. Proposed procedure to compare the local and collaborative methodologies for the fault diagnosis stage.

when the similarity measure was below the threshold $\tau = 0.90$, i.e., if $S(fdi_i(\epsilon_i)) < \tau$. The agent selection parameter σ included all twelve agent racks due to their compatibility, allowing each agent to request and collaborate with all other agents in the system. The decision-making parameter ρ was based on a majority vote, meaning the fault category most frequently diagnosed by the agents was selected.

V. EXPERIMENTAL VALIDATION AND DISCUSSION

In this section, quantitative results regarding the experiments for compare local and collaborative fault diagnosis for temperature and humidity sensors (DHT22) are presented.

A. EVALUATION METRICS

The performance metric adopted to evaluate fault diagnosis based on machine learning models is the F1-score. This metric is the harmonic mean between precision and recall, capturing their trade-off. Therefore, a higher F1-score

indicates fewer misclassifications by the fault diagnosis models.

The Wilcoxon signed-ranks test for paired samples was selected to assess whether collaboration impacts the fault diagnosis performance of twelve agents. The Wilcoxon test was selected because the experimental results did not meet the requirements for the usual T-test for two paired samples, such as normal distribution of data (checked with the Shapiro-Wilk test) and homogeneous variance (checked with the Levene test). The Wilcoxon signed-ranks test for paired samples compares whether the median difference between the samples is the same (i.e., null hypothesis) or whether the difference between the medians is different (alternative hypothesis).

B. EXPERIMENTAL RESULTS

Table 3 presents in detail the F1-Score results for data test 1 (see the left graphs of Figure 7) for the twelve agents using the different diagnostic models, considering the local and collaborative methodology based on CF*. These values (F1-score) represent the ability to classify bias, drift and precision faults that the temperature and humidity sensors are susceptible. By analyzing the mean and median values in Table3 it is evident that the collaborative approach proposed in CF* improves fault diagnosis performance across all diagnostic models compared to the local approach. A substantial example is the mean value of the agents trained with the MLP, which went from 79% to 90%, demonstrating a significant improvement in fault diagnosis performance.

When evaluating the impact of collaboration on an agent, it can be seen, for example, that the agent a7 trained with the MLP model significantly improved its performance. Without collaboration, its F1-score was 50%, but with collaboration, it reached 93%, matching the best agents. However, collaboration sometimes introduces errors, negatively affecting some agents’ performance. For instance, MLP agent a6 slightly decreased its F1-score from 90% to 89%, and SVM agent a2 dropped from 89% to 87%. These reductions may be due to suboptimal settings for the collaboration parameters (τ, σ, ρ), indicating a need for dedicated optimization to ensure collaboration benefits all agents.

TABLE 3. Result for diagnosis in terms of F1-Score (%) for data test 1.

agent	Local SVM	Collab. SVM	Local NuSVM	Collab. NuSVM	Local MLP	Collab. MLP	Local RF	Collab. RF	Local GNB	Collab. GNB
a1	81	86	79	79	89	93	81	82	46	52
a2	89	87	80	79	84	87	73	75	68	71
a3	76	77	76	78	72	93	77	79	54	56
a4	81	84	71	77	93	93	73	80	58	73
a5	81	84	77	79	81	86	90	91	74	69
a6	77	77	71	78	90	89	83	84	83	83
a7	81	83	74	78	50	93	93	91	80	81
a8	83	86	76	78	83	89	91	91	80	79
a9	84	86	79	80	83	89	66	68	61	63
a10	93	90	80	80	88	89	85	86	83	84
a11	65	71	77	81	86	93	73	74	54	63
a12	71	82	80	78	50	86	45	72	45	75
mean	80.17	82.75	76.67	78.75	79.08	90.00	77.50	81.08	65.50	70.75
median	81	84	77	79	84	89	79	81	65	72

Figure 9 compares the fault diagnosis performance between the local and collaborative methodologies for data test 1 using a box plot. Each box plot represents the performance of the twelve agents under both methodologies

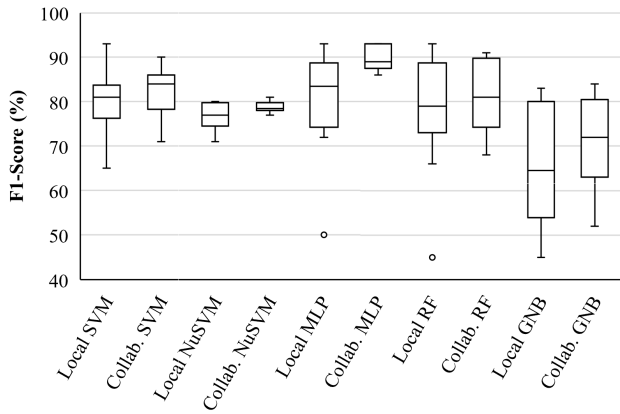


FIGURE 9. Box plot comparing the experiments based on local diagnosis with collaborative diagnosis for data test 1.

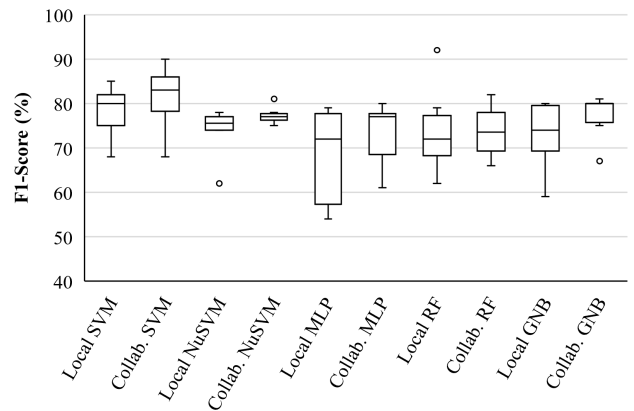


FIGURE 10. Box plot comparing the experiments based on local diagnosis with collaborative diagnosis for data test 2.

TABLE 4. Result for diagnosis in terms of F1-Score (%) for data test 2.

agent	Local SVM	Collab. SVM	Local NuSVM	Collab. NuSVM	Local MLP	Collab. MLP	Local RF	Collab. RF	Local GNB	Collab. GNB
a1	82	82	78	76	64	80	72	74	80	81
a2	82	88	74	78	56	61	67	69	73	75
a3	78	79	75	76	77	78	69	70	80	80
a4	84	86	74	77	79	80	75	78	59	80
a5	82	83	76	77	78	77	79	78	78	80
a6	71	71	62	77	74	76	78	77	73	80
a7	81	83	77	78	55	77	74	82	80	80
a8	79	90	76	77	77	77	92	92	73	80
a9	85	86	77	75	54	67	68	68	67	67
a10	68	68	75	81	61	65	72	71	68	80
a11	75	78	74	77	78	77	71	73	75	75
a12	75	83	77	77	70	73	62	66	77	78
mean	78.50	81.42	74.58	77.17	68.58	74.00	73.25	74.83	73.58	78.00
median	80	83	76	77	72	77	72	74	74	80

with different diagnostic models. The plots show that collaboration improves diagnostic capacity, as evidenced by increased minimum values for all models and a decreased interquartile range, particularly in the NuSVM, MLP, and GNB models.

To eliminate uncertainty in the results achieved, the experiments were repeated, using another test data set, the data test 2 (see this data on graphs of Figure 7). Table 4 presents the F1-Score results for each agent submitted to this test in detail. Consistent with the previous findings, both the mean and median F1-Scores were higher when agents employed collaborative behavior for fault diagnosis.

Similarly, Figure 10 compares the fault diagnosis performance between the local and collaborative methodology for data test 2 in the form of a box plot, indicating a trend of improvement in fault diagnosis when agents collaborate.

Figure 11 shows an example of the confusion matrix of an agent, in this case, selected the a8 trained with SVM model with data test 2. According to Table 4, the agent a8 had an F1-score of 79% when performing the fault diagnosis locally. However, when collaborating, the F1-Score increases to 90%, which can be confirmed with the confusion matrix in Figure 11. Specifically, collaboration notably enhanced performance in diagnosing drift faults, increasing correct classifications from 219 to 399 classifications. For bias faults, correct classifications rose from 538 to 571 classifications, and precision faults also saw improvement, with correct diagnoses increasing from 539 to 553 classifications.

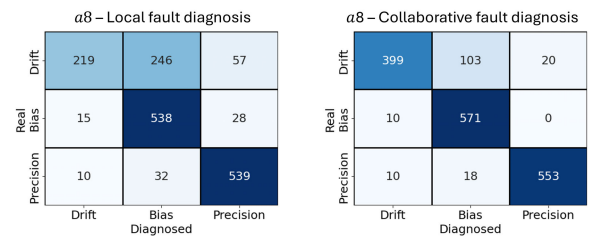


FIGURE 11. Comparison of confusion matrix of agent a8 using SVM model with data test 2.

C. WILCOXON TEST

To statistically validate the impact of collaboration for the experiments presented, the Wilcoxon signed-ranks test for paired samples was performed to verify whether there is a significant difference between the results obtained locally and with collaboration. For more details of this statistical test, it is recommended to study the Chapter 18.6 and 18.7 of [33].

Thus, according to the Wilcoxon test, the following null hypothesis H_0 and alternative hypothesis H_1 are established respectively:

- H_0 : The distribution of the medians of the F1-Score differences for fault diagnosis in the population (i.e., local vs. collaborative agent) is symmetric close to zero, i.e., indicating that any difference is due to chance.
- H_1 : The distribution of the median F1-Score differences for local fault diagnosis in the population is smaller than the distribution of the median F1-Score differences for collaborative fault diagnosis in the population of agents.

The test results are presented in Table 5. The values of T^+ and T^- are statistics calculated for the Wilcoxon test from the difference between the pairs of samples (i.e., local and collaborative agent), where T^+ is the sum of the positive differences, and T^- is the sum of the negative differences. The statistic value W is the value used in the Wilcoxon test where the smaller between T^+ or T^- is selected. The p-value is obtained from the W statistic, and for $p - value < 0.05$ the null hypothesis H_0 is rejected with 95% of confidence.

TABLE 5. Result of Wilcoxon Signed-Rank for paired samples test.

	Model	T ⁺	T ⁻	W	p-value	Significance
data test 1	SVM	9,5	56,5	9,5	0,0161	reject H_0
	NuSVM	9	46	9	0,0137	reject H_0
	MLP	1,5	64,5	1,5	0,0010	reject H_0
	RF	7,5	58,5	7,5	0,0093	reject H_0
	GNB	15	21	15	0,0181	reject H_0
data test 2	SVM	0	45	0	0,0020	reject H_0
	NuSVM	11	55	11	0,0269	reject H_0
	MLP	5	61	5	0,0049	reject H_0
	RF	7,5	47,5	7,5	0,0186	reject H_0
	GNB	0	36	0	0,0039	reject H_0

Therefore, according to Table 5, the fault diagnosis performance, in terms of F1-Score, for agents without collaboration (i.e. local) was lower than the performance of collaborative agents adopting CF* methodology. The Wilcoxon signed-rank test for paired samples showed that this difference is statistically significant for the five diagnosis models tested considering the two different data test.

VI. CONCLUSION AND FUTURE WORKS

This work introduces a collaborative fault diagnosis approach for cyber-physical systems based on the CF* methodology, leveraging multi-agent system concepts. The agent behavior was parameterized to define when to initiate collaboration, which agents to collaborate with, and how made decision based on the collaboration to enhance fault diagnosis. Moreover, collaboration protocols inspired by FIPA-ACL were implemented. Consequently, intelligent and autonomous agents utilize information exchange and collaboration to transcend their individual capabilities in fault diagnosis.

A case study was conducted using twelve racks in a small-scale laboratory warehouse to test the fault diagnosis capabilities of DHT22 sensors with injected shift, bias, and precision faults. The experiments focused on the diagnosis stage, comparing local and collaborative methodologies. Each agent was trained using five machine-learning methods and subjected to two data tests. The results consistently showed that collaborative fault diagnosis improved performance in median by up to 10% compared to the traditional local methodology. Additionally, the statistical Wilcoxon test confirmed that the collaborative approach significantly outperformed the local methodology in all experiments. The proposed CF* methodology is suitable for others CPS application.

This paper recognizes some limitations of CF*, primarily in addressing the interoperability challenge among diverse CPS agents. The complexity of ensuring seamless integration and communication across agents with varying vendors, protocols, and operational models can limit collaborative efficiency for fault diagnosis. Another challenge lies in developing collaboration models for heterogeneous agents (e.g., collaboration between agents representing storage rack and robot), where agents representing different resources or incompatible systems must effectively cooperate and contribute uniquely to fault diagnosis. Ensuring meaningful and coordinated interactions across such varied agents

requires advanced alignment strategies to maximize diagnostic effectiveness and maintain system resilience.

In summary, this research paves the way for further studies in collaborative fault tolerance for CPS and lays the groundwork for future advancements. Future work will focus on developing a digital twin to improve the diagnosis and collaboration models of CF* dynamically, thereby enhancing fault tolerance performance. Additionally, future work will explore and modeling real-time performance impact of the collaboration for fault diagnosis in CF*, which is crucial for safety-critical CPS applications, where collaboration time may affect fault tolerance by requiring real-time fault diagnosis results.

ACKNOWLEDGMENT

This work was supported by the Coordination for the Improvement of Higher Education Personnel-CAPES under Grant ROR identifier: 00x0ma614 for the Article Processing Charge for the publication of this research. This work was supported in part by Fundação para a Ciência e a Tecnologia (FCT) through Portugal National funds FCT/MCTES (PIDDAC): CeDRI under Grant UIDB/05757/2020 (DOI: 10.54499/UIDB/05757/2020) and Grant UIDP/05757/2020 (DOI: 10.54499/UIDP/05757/2020), and in part by SusTEC under Grant LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020). The work of Luis Piardi was supported by FCT for the Ph.D. under Grant UI/BD/151286/2021 (DOI: 10.54499/UI/BD/151286/2021).

REFERENCES

- [1] I. L. D. Makanda, P. Jiang, M. Yang, and H. Shi, "Emergence of collective intelligence in industrial cyber-physical-social systems for collaborative task allocation and defect detection," *Comput. Ind.*, vol. 152, Nov. 2023, Art. no. 104006.
- [2] R. Baheti and H. Gill, "Cyber-physical systems," *Impact Control Technol.*, vol. 12, no. 1, pp. 161–166, 2011.
- [3] Y. Liu, Y. Peng, B. Wang, S. Yao, and Z. Liu, "Review on cyber-physical systems," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 1, pp. 27–40, Jan. 2017.
- [4] K. Zhang, Y. Shi, S. Karnouskos, T. Sauter, H. Fang, and A. W. Colombo, "Advancements in industrial cyber-physical systems: An overview and perspectives," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 716–729, Jan. 2023.
- [5] Y. Lu, "Industry 4.0: A survey on technologies, applications and open research issues," *J. Ind. Inf. Integr.*, vol. 6, pp. 1–10, Jun. 2017.
- [6] F. Januário, A. Cardoso, and P. Gil, "A distributed multi-agent framework for resilience enhancement in cyber-physical systems," *IEEE Access*, vol. 7, pp. 31342–31357, 2019.
- [7] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.
- [8] L. Piardi, P. Costa, A. Oliveira, and P. Leitão, "Collaborative fault detection and diagnosis architecture for industrial cyber-physical systems," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Aug. 2022, pp. 1–6.
- [9] J. Lee, E. Lapira, B. Bagheri, and H.-A. Kao, "Recent advances and trends in predictive manufacturing systems in big data environment," *Manuf. Lett.*, vol. 1, no. 1, pp. 38–41, Oct. 2013.
- [10] B. Dafflon, N. Moalla, and Y. Ouzrout, "The challenges, approaches, and used techniques of CPS for manufacturing in Industry 4.0: A literature review," *Int. J. Adv. Manuf. Technol.*, vol. 113, nos. 7–8, pp. 2395–2412, Apr. 2021.
- [11] L. Piardi, P. Leitão, and A. S. de Oliveira, "Fault-tolerance in cyber-physical systems: Literature review and challenges," in *Proc. IEEE 18th Int. Conf. Ind. Informat. (INDIN)*, vol. 1, Jul. 2020, pp. 29–34.

- [12] R. Isermann, *Fault-Diagnosis Systems: An Introduction From Fault Detection to Fault Tolerance*. Cham, Switzerland: Springer, 2006.
- [13] F. Rojas, C. Jerez, C. M. Hackl, O. Kalmbach, J. Pereda, and J. Lillo, "Faults in modular multilevel cascade converters—Part II: Fault tolerance, fault detection and diagnosis, and system reconfiguration," *IEEE Open J. Ind. Electron. Soc.*, vol. 3, pp. 594–614, 2022.
- [14] B. Dowdeswell, R. Sinha, and S. G. MacDonell, "Finding faults: A scoping study of fault diagnostics for industrial cyber-physical systems," *J. Syst. Softw.*, vol. 168, Oct. 2020, Art. no. 110638.
- [15] H. Zheng, R. Wang, Y. Yang, J. Yin, Y. Li, Y. Li, and M. Xu, "Cross-domain fault diagnosis using knowledge transfer strategy: A review," *IEEE Access*, vol. 7, pp. 129260–129290, 2019.
- [16] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on long short-term memory networks," *Renew. Energy*, vol. 133, pp. 422–432, Apr. 2019.
- [17] L. M. R. Baccarini, V. V. Rocha e Silva, B. R. de Menezes, and W. M. Caminhas, "SVM practical industrial application for mechanical faults diagnostic," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6980–6984, Jun. 2011.
- [18] B. Xia, K. Wang, A. Xu, P. Zeng, N. Yang, and B. Li, "Intelligent fault diagnosis for bearings of industrial robot joints under varying working conditions based on deep adversarial domain adaptation," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.
- [19] Y. Seng Ng and R. Srinivasan, "Multi-agent based collaborative fault detection and identification in chemical processes," *Eng. Appl. Artif. Intell.*, vol. 23, no. 6, pp. 934–949, Sep. 2010.
- [20] F. Zeng, Z. Li, Z. Zhou, and S. Du, "Fault classification decision fusion system based on combination weights and an improved voting method," *Processes*, vol. 7, no. 11, p. 783, Nov. 2019.
- [21] H. Shao, J. Lin, L. Zhang, D. Galar, and U. Kumar, "A novel approach of multisensory fusion to collaborative fault diagnosis in maintenance," *Inf. Fusion*, vol. 74, pp. 65–76, Oct. 2021.
- [22] N. Ali, M. Hussain, and J.-E. Hong, "Analyzing safety of collaborative cyber-physical systems considering variability," *IEEE Access*, vol. 8, pp. 162701–162713, 2020.
- [23] N. Ali, M. Hussain, and J.-E. Hong, "Fault-tolerance by resilient state transition for collaborative cyber-physical systems," *Mathematics*, vol. 9, no. 22, p. 2851, Nov. 2021.
- [24] L. Piardi, A. S. de Oliveira, P. Costa, and P. Leitão, "Collaborative fault tolerance for cyber-physical systems: The detection stage," *Comput. Ind.*, vol. 166, Apr. 2025, Art. no. 104253.
- [25] L. Piardi, A. Oliveira, P. Costa, and P. Leitão, "Evaluation metrics for collaborative fault detection and diagnosis in cyber-physical systems," in *Proc. IEEE 29th Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2024, pp. 1–8.
- [26] S. Poslad, "Specifying protocols for multi-agent systems interaction," *ACM Trans. Auto. Adapt. Syst.*, vol. 2, no. 4, p. 15, Nov. 2007.
- [27] L. Piardi, P. Costa, A. Oliveira, and P. Leitão, "MAS-based distributed cyber-physical system in smart warehouse," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 6376–6381, 2023.
- [28] S. U. Jan, Y. D. Lee, and I. S. Koo, "A distributed sensor-fault detection and diagnosis framework using machine learning," *Inf. Sci.*, vol. 547, pp. 777–796, Feb. 2021.
- [29] S. Zidi, T. Moulahi, and B. Alaya, "Fault detection in wireless sensor networks through SVM classifier," *IEEE Sensors J.*, vol. 18, no. 1, pp. 340–347, Jan. 2018.
- [30] S. U. Jan, Y.-D. Lee, J. Shin, and I. Koo, "Sensor fault classification based on support vector machine and statistical time-domain features," *IEEE Access*, vol. 5, pp. 8682–8690, 2017.
- [31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. J. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2012.
- [32] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proc. 22nd Int. Conf. Mach. Learn. (ICML)*, 2005, pp. 625–632.
- [33] D. C. Howell, *Statistical Methods for Psychology*. Boston, MA, USA: PWS-Kent Publishing Co, 1992.



LUIS PIARDI received the degree in electronic engineering from Universidade Tecnológica Federal do Paraná (UTFPR), Brazil, in 2018, and the master's degree in industrial engineering from Instituto Politécnico de Bragança, Portugal, in 2018. He is currently pursuing the Ph.D. degree in electrical and computer engineering through the Cotutelle program with UTFPR, and the Faculty of Engineering, University of Porto (FEUP), Portugal. Since 2018, he has been a Researcher with the Research Centre in Digitalization and Intelligent Robotics (CeDRI). His research interests include fault-tolerance, industrial cyber-physical systems, intelligent production systems, the IoT, and multi-agent systems.



PEDRO COSTA received the M.Sc. degree from the Electrical Engineering Department, Faculty of Engineering, University of Porto (FEUP), in 1999, and the Ph.D. degree in electrical and computer engineering from FEUP in the area of control and robotics, with the thesis "Planning Cooperative tasks and trajectories in Multiple Robots," in 2011. He is currently a Professor with the Computers and Electrical Engineering Department, FEUP, and a Researcher in robotic and intelligent systems with the Institute for Systems and Computer Engineering, Technology, and Science (INESC TEC), Porto. His research interests include robotics, automation, and cyber-physical systems.



ANDRÉ SCHNEIDER DE OLIVEIRA (Member, IEEE) received the M.Sc. degree in mechanical engineering, concentrate on force control of rigid manipulators, from the Federal University of Santa Catarina (UFSC), in 2007, and the Ph.D. degree in engineering of automation and systems, in 2011, with the thesis focused on differential kinematics through dual-quaternions for vehicle-manipulator systems. He is currently an Associate Professor with Universidade Tecnológica Federal do Paraná (UTFPR) and a member of the Advanced Laboratory of Robotics and Embedded Systems (LASER) and the Laboratory of Automation and Advanced Control Systems (LASCA). His research interests include robotics, mechatronics, and cyber-physical systems.



PAULO LEITÃO (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Porto, Portugal, in 2004. He is currently a Full Professor with the Department of Electrical Engineering, Instituto Politécnico de Bragança, Bragança, Portugal, and the Coordinator of the Research Centre in Digitalization and Intelligent Robotics (CeDRI). His research interests include intelligent and reconfigurable systems, industrial cyber-physical systems, multi-agent systems, digital twin, and factory automation. He is a Senior Member of the IEEE Industrial Electronics Society (IES) and the Systems, Man, and Cybernetics Society (SMCS), and the Chair of the IEEE IES Technical Committee on Industrial Agents and the Established IEEE 2660.1-2020 Standard.

• • •